# 意味表現の解明に向けて：ベクトル空間モデルのアナロジー推論の分析
# Towards Understanding of the Semantic Representation: Analysis of Analogical Reasoning by Vector Space Models

加藤龍彦 [1)]，日高昇平 [1)]
KATO Tatsuhiko [1)]，HIDAKA Shohei[1)]
skylark@jaist.ac.jp, shhidaka@jaist.ac.jp

1) 北陸先端科学技術大学院大学
1) Japan Advanced Institute of Science and Technology,

【要約】本稿では Mikolov ら (2013a) によって提案された Skip-gram アルゴリズムの探索的な検討の結果を報告する．我々は以下の二点を分析した．(1) Skip-gram によって生成された単語の低次元空間での可視化 (2) Skip-gram の単語表現とデータセットの単語共起確率の行列による表現とのアナロジー課題を用いた比較． 我々は以上の分析結果が，Skip-gram モデルがアナロジー課題において扱うのを不得意とするようなデータ中の関係性があること，またアナロジー課題の正答にとって重要な特性は，Skip-gram モデルの特性そのものよりもデータセット中の共起関係にあることの二点を示唆すると主張する．

【キーワード】意味表現　アナロジー　機械学習

## 1. はじめに

本稿では意味表現という用語で，言葉や物事の間に見出された関係性を，記憶に保持しておく際に利用する形式を指す．このような意味表現を作り出す能力は，人の認知の働きを理解する上で本質的である．推論，記憶，言語といった認知科学の古典的な問題はすべて意味表現と関わっている．推論を行う際には適切な因果関係の知識を選択してそれを問題に当てはめる必要があり，その際には事象と因果の関係を理解している必要がある．物事の関係性を利用する連想は記憶の基本的な性質である．またある種のアナロジー推論を行うにはさらに進んで，関係性同士の関係性を見出す必要もある (Holyak & Thugaard, 1994)．意味表現の問題とは，こうした認知活動を行う上で人がどのような表現を用いているのかという問いである．

本研究は，アナロジーを通して人の意味表現の特性を定式化することを目的とする．アナロジーを作るには対象の間の関係から意味表現を作り，その表現を用いて異なる対象間に関係を見出す，という作業が必要になる．ここには意味表現の生成とその使用に関わる問題がどちらも含まれている．適切なアナロジー推論を行うには何らかの意味表現を作っておいてそれらの間の関係を見出す必要があるし，アナロジーを使用して他の事象を想像する時にはすでにある意味表現を使用して異なる事象を考え出している．そのため，我々はアナロジーの研究を通じて意味表現の特性を理解する，という立場を取る．

意味表現のモデリング研究は認知科学において早くから行われてきたが (e.g. Collins and Quillian, 1969; Collins and Loftus, 1975)，特にアナロジーに関して興味深い最近の結果として，自然言語処理の分野において提案された Skip-gram モデルや GloVe モデルがある．(Mikolov et al., 2013a, b; Pennington et al., 2014)．これらのモデルはベクトル空間モデルと呼ばれ，比例アナロジー課題 (“A is to B as C is to ?”という文において，単語 A, B, C の関係から D を予測する課題) に対して高い精度を持つことが確かめられている．しかし一方で，こうしたモデルは人のように対象の属性を参照して課題をこなしたり，新奇なアナロジーに命名する，といったことはできない．本研究はこのような点を踏まえて，Skip-gram モデルに代表されるベクトル空間モデルの分析を通じて人の意味表現の性質を明らかにすることを目的とする．

本稿ではまず，２節において Skip-gram のアルゴリズムと Mikolov らの行ったアナロジー課題について概観し，彼らのアルゴリズムが従来のニューラルネットワークモデルに比べて極めて単純化されたものであることを確認する．その後我々の分析結果を紹介する．概要としては，簡単な単語の共起確率の行列によってのみでも小さなデータセットについては，Mikolov らのモデルに近いアナロジー課題の精度が出ることを確かめた．もちろん計算量に関しては Mikolov らの計算モデルの方が遥かに効率的だが，

我々の分析結果は Skip-gram のアナロジー課題における高い予測精度が特定のモデルの特性によるものではなく，データセットの共起関係そのものの特性によることを示唆する．

## 2. 意味の分布仮説と単語共起モデル

機械翻訳や質問回答システムなどの言語処理を行なう際にまず問題となるのは，語をどのように表現するか，という問題である．その場合最も単純な方法は，各語を離散的に表現することである．このような表現はワンホットベクトルと呼ばれ，データ中における出現語のインデクスに対応するベクトルの次元の値を 1 とし，他を 0 とすることで表現する．しかしこのような表現では，文書中にある語が出現するかどうかは分かっても，単語間の関係性まで捉えることはできない．なぜなら各単語がベクトルのある要素を 1 にすることで表現されているため，どのベクトル間の距離も等しいからである．

このような理由から，語のいわゆる「分散表現」を考えることが重要になる．分散表現とは，各語を連続的な確率値を有するベクトルとして表現する方法であり，通常ある語とその周辺の文脈語との共起関係を表現する．このような単語間の共起関係を重視する根拠として頻繁に参照されるのは，Harris (1954) の提唱した *Distributional Hypothesis* である．それによれば，「…各言語は分布構造，つまり，ある部分の他の部分との相対的な出現頻度という観点から説明可能である．」(Harris, 1954)，あるいは「ある語の意味はその周辺語から知ることができる．」 (Firth, 1957)．こうした語の共起関係を重視する発想は Mikolov らのモデルにも強く現れている．

## 3. Skip-gram のアルゴリズムと我々の仮説

それでは Skip-gram のアルゴリズムを簡単に述べよう．Skip-gram のアルゴリズムは，ある語からその周辺語への予測を最大にするような単語表現を生成する，という発想の下に作られており，基本的なアーキテクチャは以下のようになっている (図 1).

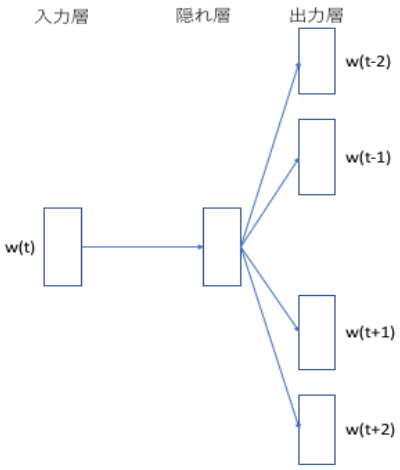


図 1: Skip-gram モデルのアーキテクチャ

ここで w(t)は *t* 番目の語であり，w(t-i)から w(t+i)までの語の予測を最大化するようにベクトル化する．つまり，*t* 番目の語が与えられた時の周りの i 語の対数条件付き確率の平均，

$$\frac{1}{T}\sum\sum \log p(w_{t+i} \mid w_t)$$

を最大化する．Mikolov らは語$w_I$が与えられた時の，語$w_O$の条件付き確率を以下のように語のベクトル$v_{w_I}, v_{w_O}$の関数として定義した．

$$p(w_O \mid w_I) = \frac{\exp\left(v_{w_O}^{\top} v_{w_I}\right)}{\sum_{w=1}^{W} \exp\left(v_{w_O}^{\top} v_{w_I}\right)}$$

実際の Skip-gram では計算量の効率化のために上記のソフトマックス関数を単純化した階層的ソフトマックス法やネガティブ・サンプリング法，高頻度で出現する語を一定確率で捨てるサブサンプリング法を用いたが，本稿での議論には関わらないので割愛する．

我々の議論にとって重要な点は，Skip-gram のアルゴリズムの単純さにある．上述のように，Skip-gram

のアルゴリズムは単層のニューラルネットワークであり，これにソフトマックス関数を適用しているにすぎない．それにも関わらず，このアルゴリズムは意味，統語的な課題に対して既存のモデルを大きく上回るパフォーマンスを発揮する．

Skip-gram で扱われているアナロジー課題は，前述した比例アナロジーである．この課題を解くにあたって Skip-gram では，$v_i$をそれぞれ語 $i$ のベクトル表現だとすると，3 つの語 $a, b, c$ に関して $v_a - v_c + v_b$ というベクトル演算を行い，その結果得られたベクトルに最も類似する$v_d$を答えとして選択する，という手法を取る．以上の操作を，$v_a - v_c + v_b \approx v_d$と表記する．よく知られた例で言えば，$v_{king} - v_{man} + v_{woman} \approx v_{queen}$ という結果になればモデルは正しく課題に回答した，ということになる．このような演算を図示すると，$v_{king} - v_{man}$と$v_{queen} - v_{woman}$が以下の図 2 ような平行関係にある時に演算結果が正しいアナロジーとなる．

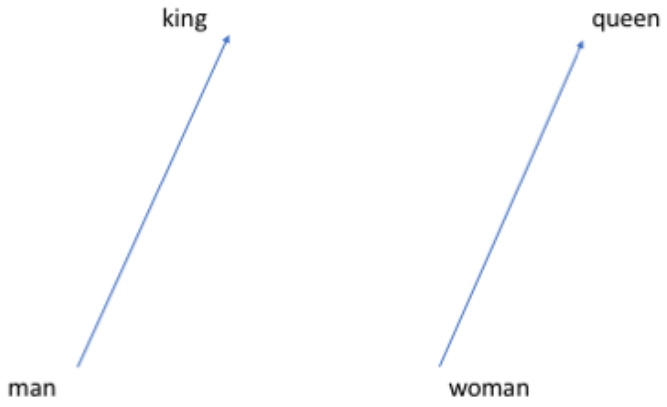


図 2: ベクトル演算による比例アナロジーの成立条件は 2 つの単語対の間に平行関係が成り立っていること

Mikolov ら (2013a) の報告するところでは，通常の非線形変換を入れたようなニューラルネットワーク言語モデルの意味課題に対する正答率は 23%，統語課題に関する正答率は 53%であるのに対して，Skip-gram の正答率はそれぞれ 55%，59%である．このことは，Mikolov らのタスクで扱われているような意味，統語関係について Skip-gram モデルはそれまでのモデルよりも優れた表現を形成していることを示唆している．では，Skip-gram がその単純な構造にも関わらず言語の重要な構造を取り出し得るのはなぜなのだろうか．我々はこのような目的意識の下で Skip-gram に関して探索的な分析を行った．

## 4. Skip-gram の分析と単語共起行列のパフォーマンス

以下の我々の分析では，まず Skip-gram の単語ベクトルの低次元における可視化を行い，Skip-gram においてアナロジーが成立しているような単語の組が，そのモデルの対象となったデータセットにおいては共起関係の比率が一定であることを確認した．その後より直接的な検証を行うため，Skip-gram とデータの共起行列を直接的に計算するモデルを Brown corpus (Francis & Kucera, 1979) について訓練した．そしてこれらの単語ベクトルを用いて，上記の Mikolov ら (2013a) が行っているのと同様のアナロジー課題を行わせた．すると，データから直接に共起行列を計算するようなモデルでもアナロジー課題において Skip-gram と同程度の正答率でアナロジー課題をこなし得ることが示された．

まずは低次元における可視化の結果を示す．データセットとしては Google が公開している，Google News から取得したおよそ 1000 億語について訓練したものを用いた．この訓練済みデータは語と句計 300 万を含み，それぞれの語句が 300 次元を用いて表現されている(1) (Mikolov ら (2013a) で示されている分析結果は 1000 次元ベクトルによるもの)．図 3 はこのモデルについていくつかのクラスの語の主成分分析を行い，300 次元空間のある 2 次元部分空間について可視化を行ったものである．左上の国と首都に関する図は Mikolov ら (2013a) の図を再現したものだが，いくつかの国と首都の間にアナロジー課題に正解するために必要な，前述したような平行関係が成立していることがわかる．これは男女関係についても同様で，それに加えて「家族」，「職業」，「歴史」，と言えるようなクラスターが発生していることも分かる．しかし例えば主語-動詞関係について見てみると，I-am, He-is, She-is などの間では

アナロジー関係が成立していない．このことは，$v_a + v_b - v_c \approx v_d$ という関係式では捉えきることができないような関係が存在することを示唆する．

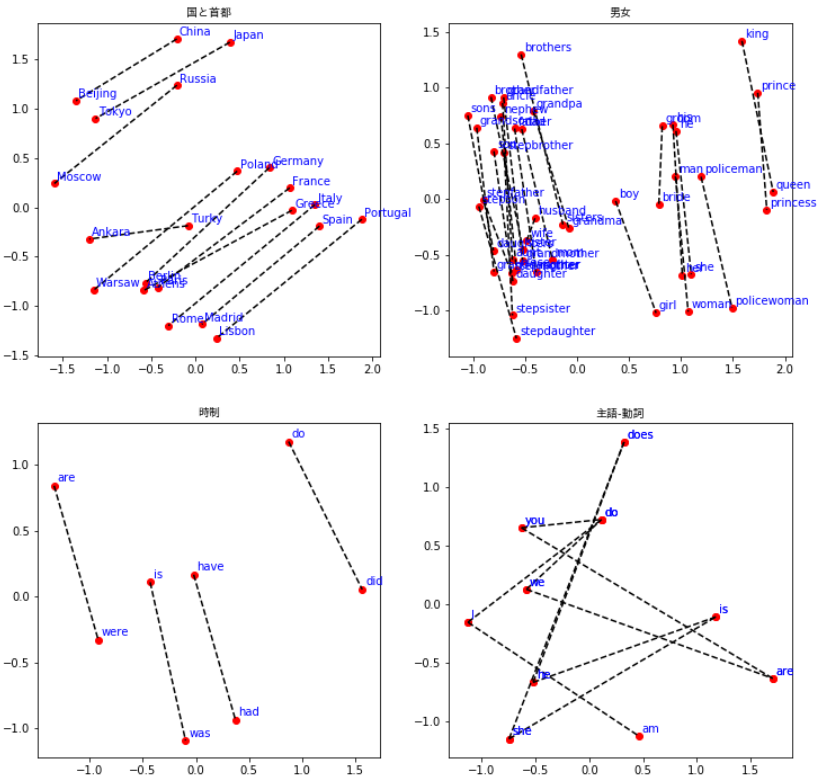


図 4： Skip-gram 中の単語ベクトルの幾つかを主成分分析によって 2 次元空間に射影したもの．国と首都，男女，時制関係では平行関係(図 2)が成立しているが，主体-動詞関係を示した図では平行関係が一部でしか成立していない．

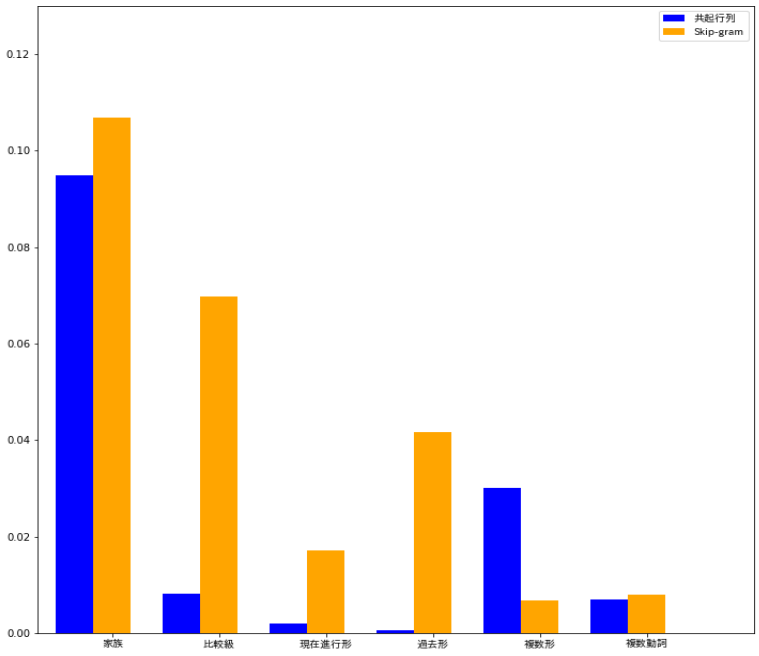


図 3 : 2 つの単語表現のアナロジー課題に関する正答率の比較．データセットは Mikolov ら (2013a) が用いたものから「家族」，「比較級」，「現在進行形」，「過去形」，「複数形」，「複数動詞」カテゴリーを選択した．

次に，我々は Brown corpus を用いて単語の共起行列を作成し，同じコーパスで Skip-gram モデルを訓練した．その上で 2 つのモデルに関して，Mikolov ら (2013a) が用いているアナロジー課題の一部に関する正答率を比較した．元のデータセットは 14 のカテゴリーからなり全体で 19544 の項目から構成されるが，本稿ではデータセット中での対象語の頻度が低いために比較できなかったいくつかのカテゴリについては除外し，残った 6656 項目について分析した．その結果は図 4 に示している．ここから見て取れるように，Brown corpus のような比較的小さいコーパスについては Skip-gram でも，データセット中の共起関係をほとんどそのまま用いたような表現でも，アナロジー課題における精度に大きな差異は見られなかった．

## 5. 議論と予想

4 節までで我々が行った探索的な分析において以下の二点の結果が得られた．

1. Skip-gram モデルによるベクトル表現には，比例アナロジーを平行なベクトルの関係式で表せる単語とそうでないものが混在する．

2. Skip-gram を用いて作成された単語表現と元のデータセットの共起確率分布をそのもののベクトル表現との間では，比例アナロジー課題の予測精度の差異がほとんど見られない．

1 の結果は, データ中の関係に単なる平行関係では捉えられないものがあることを意味する. しかし，Mikolov らは平行関係によってのみ 4 項の比例アナロジーを検討しており，それ以外の関係を持つアナロジーの抽出に関する Skip-gram モデルの有用性には疑問が残る．この点はより詳しく検討する余地があるだろう．

2 の結果は，アナロジー課題で予測を行うために重要なポイントは Skip-gram のアーキテクチャ自体にあるわけではなく，コーパスデータ中に潜在する共起構造にあることを示唆している．我々は単純に語の共起頻度の確率を単語ベクトルとして表現しており，Skip-gram のようなモデル化や近似を一切行っていない．しかし，こうしたデータの確率ベクトルを使って，アナロジー課題に関して Skip-gram に比肩するような精度を出すことができることを示した．これは，(Skip-gram によるベクトル表現の構築ではなく，)語の共起頻度そのものが比例アナロジー課題に有用な構造を潜在的に有しているためだと考えられる．

## 6. 結論と展望

本稿では，Skip-gram のアルゴリズムを記述し，我々が行った Skip-gram の探索的な検討から得られた 2 つの結果とその解釈を示した．1 つは，データ中には Skip-gram が用いているような線形のベクトル関係だけでは捉えられないような構造が存在するという点，もう 1 つは比例アナロジー課題を解く際に必要となる関係の表現を作るにあたって，必ずしも Skip-gram のアーキテクチャが特別な役割を果たしているわけではなく，語の共起関係がこうした関係に関する情報を多く含んでいる可能性がある，という点である．

我々はこれらの結果を踏まえて，語の共起関係を用いて関係を発見するアルゴリズムを開発する予定である．今回の 2 つ目の結果から示唆されるように，語に関する本質的な情報は，言語における語の分布中の共起関係が有していると考えられる．そうであれば，この共起関係に関する情報から直接的に比例アナロジーの予測に有用な性質を抽出するようなアルゴリズムを作れば，多くの先行研究で仮定されているような平行関係に限らない，より多種の比例アナロジーにも対応できると考えられる．このようなアルゴリズムは，本稿の結果 1 が示唆するような，平行関係の他にどのような関係がアナロジー課題を解くのに使用可能か，といった問にも回答することができると考えられ，比例アナロジーの一般的な理論となり得る可能性を持つ．

## 注

(1) https://code.google.com/archive/p/word2vec/

## 参考文献

Collins, A. M., & Loftus, E. F., (1975). A Spreading-Activation Theory of Semantic Processing. Psychological Reviews, Vol. 82, No.6, 407-428.

Collins, A. M., & Quillian, M. R. (1969) Retrieval Time from Semantic Memory. Journal of Verbal Learning and Verbal

Behavior, 8, 240-248.
Francis, W. N, and Kucera, H. A Standard Corpus of Present-Day Edited American English, for use with Digital Computers. (1979). Brown University. Providence, Rhode Island.
Holyoak, K. J., & Thagard, P. (1996). Mental Leaps. MIT Press. (ホリオーク, K. J., & サガード, P. (1998) 『アナロジーの力: 認知科学の新しい探求』 (鈴木宏明・河原哲雄訳) 新曜社)
Mikolov, T., Chen, K., Corrado, G., and Dean, J., (2013a). Efficient Estimation of Word Representations in Vector Space. In ICLR Workshop Papers.
Mikolov, T., Sutskever, I., Chen, K., Corrado, K., and Dean, J., (2013b). Distributed Representations of Words and Phrases and Their Compositionality. In NIPS, pages 3111‐3119.
Pennington, J., Socher, R., Manning, C., D., (2014). GloVe: Global Vectors for Word Representation. Retrieved from https://nlp.stanford.edu/pubs/glove.pdf
Sahlgren, M. (2008). The distributional hypothesis. Italian Journal of Disability Studies, 20, 33-53.

連絡先
住所：〒923-1211　石川県能美市旭台 1-8-7-403
名前：加藤龍彦
E-mail：skylark@jaist.ac.jp